International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 3 Issue 5, August 2019 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Voice Command Mobile Phone Dialer

Aye Thida, Yee Wai Khaing

Artificial Intelligence Lab, University of Computer Studies, Mandalay, Myanmar

How to cite this paper: Aye Thida | Yee Wai Khaing "Voice Command Mobile Phone Dialer" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-3 | Issue-5, August 2019, pp.1904-1909, **https://doi.org/10.31142/ijtsrd26814**

IJTSRD26814

ABSTRACT

Speech recognition is an advanced technology that uses desired equipment and a service which can be controlled through voice without touching the screen of the smart phone. In current century, there are many researches with the help of speech recognition on mobile devices. In this system, mobile phone users can command with their voice to easily make phone call. Google's cloud speech API is used to recognize the incoming user voice. The speech API recognizes over 120 languages but it cannot correctly provide Myanmar Language still now. The system will classify the Myanmar proper name recognized by the Google's speech API to get the correct name with the help of Naïve Bayesian Classifier. The contact name classified by Naive Bayes can only meet user's desired one just written in English script and it cannot provide the name written in Myanmar script. This system uses hybrid transliteration approach to solve the contact name recorded by Myanmar script. Therefore the system can make phone call to the contact name typed with not only English script but also Myanmar script. The system applies Jaro-Winker distance measure to outperform the accuracy of system output. Success rate is used to measure the performance of each process contained in the system. This system is implemented with Android programming language.

KEYWORDS: *speech recognigiotn; voice command; mobile phone dialer; styling;*

I. INTRODUCTION

Speech is the easiest and most common way for people to communicate. Speech is also faster than typing on a keypad and more expressive than clicking on a menu item. Therefore speech applications that recognize the human's natural voice are becoming more popular in human life. From the research point of view, there are many researches with the help of speech recognition. Generally the meaning of speech recognition is that it is an advanced technology that translates spoken words into text. Speech recognition is one of the NLP's major tasks. Some NLP's tasks are information retrieval, question and answering, machine translation, speech segmentation, transliteration and so on. Nowadays there are plenty of applications and areas where speech recognition is used. Most mobile internet devices are making interesting use of speech recognition. The iPhone and Android devices are examples of that. Examples of mobile applications, which implement recognize user's speech in these smartphone devices, are Siri (only iPhone) and Google Now (both iPhone and Android). In this system, it will implement phone calling with speech on android mobile devices with the help of Google's speech recognition engine. The speech API can correctly recognize English spoken words but not Myanmar language words. For the speech input of Myanmar words, it can only recognize English words which pronunciation is similar to the input Myanmar word. Classification andsimilarity methods are needed in order to correctly recognize Myanmar spoken words.

There are different kinds of systems that use Google speech recognition engine or Google Speech API. Some of these systems are as following:

B. Raghavendhar Reddy, E. Mahender [2] proposed an application for sending SMS messages which uses Google's speech recognition on. The Voice SMS application allows user to input spoken information and send voice message as desired text message. The user is able to manipulate text message fast and easy without using keyboard, reducing spent time and effort. In this case speech recognition provides alternative to standard use of key board for text input, creating another dimension in modern communications.

Sagarjit Dash [3] introduced voice detection capability in the quiz application for Android platform smart mobile phones with the help of Google Speech API. Here the device proposed is an interactive android smart phone, which is capable of recognizing spoken words. He proposed to develop interactive application which can run on the tablet or any android based phone. The application helps the user to give the answer of a question using voice and through voice user can go to the next question. Users can command a mobile device to do something via speech. These commands are then immediately executed.

Ms. Anuja Jadhav Prof. Arvind Patil [4] demonstrated android speech to text converter for SMS application and also the requirement of speech to text conversion system. This converter based on evaluating voice versus keypad as a means for entry and editing of texts. In other words, mobile SMS users can say messages can be voice/speech typed. A speech to text converter is developed to send SMS. It is found that large-vocabulary speech recognition can offer a very competitive alternative to traditional text entry. This SMS

application also implements speech recognition process at Google's speech server.

Hae-Duck J. Jeong, Sang-Kug Ye, Jiyoung Lim, Ilsun You, and WooSeok Hyun [5] had proposed a computer remote control system using voice recognition technologies of mobile devices and wireless communication technologies for the blind and physically disabled population as assistive technology. These people experience difficulty and inconvenience using computers through a keyboard and/or mouse. The purpose of this system is to provide a way that the blind and physically disabled population can easily control many functions of a computer via voice. The configuration of the system consists of a mobile device such as a smartphone, a PC server, and a Google server that are connected to each other. Users can command a mobile device to do something via voice; such aswriting emails, checking the weather forecast, or managing a schedule. These commands are then immediately executed.

The proposed system also provided blind people with a function via TTS (Text to Voice) of the Google server if they want to receive contents of a document stored in a computer. People are interested in mobile phones because they can actually stay in touch wherever they are. Now multi-touch surface have been widely used as the main input methods on mobile devices. But for visually impaired, it is difficult to use these methods. Moreover nowadays people are very busy and so they do their works in a timely fashion. For example, at the time of driving, a user wants to phone a call in his or her contact list and so he or she searches it in many contacts. Consequently it makes wasting time and even unexpected accidents may occur. Therefore a speech recognition application for mobile device is being developed to avoid harmful accidents and it can save user's valuable time. Moreover voice control is an effective and efficient alternative non-visual interaction mode which does not require target locating and pointing.

II. Speech Recogntion

Speech recognition is the ability of a machine or program to identify words and phrases in spoken language and convert them to a machine-readable format. Speech recognition, also known as speech-to-text or automatic speech recognition (ASR) has been studied for more than two decades and recently been used in various commercial products. There are plenty of applications and areas where speech recognition is used. This technology has been more popular and successful in the following:

- Device control. For individuals with disabilities and birth defects that leave them unable to use their hands, NLP technology can be used to control a mobile device. For example, just saying "OK Google" to an Android phone fires up a system that is all ears to your voice commands.
- Car Bluetooth systems. Many cars are equipped with a system that connects its radio mechanism to the smartphone through Bluetooth. Phone calls can be made and received without touching the smartphone, and can even dial numbers by just saying them.
- Voice transcription. In areas where people have to type a lot, some intelligent software captures their spoken words and transcribes them into text. This is current in certain word processing software.

III. Architecture of the System

When the system starts, it transliterates Myanmar contact name that is recorded by Myanmar script in the contact list to English script. And the system recognizes user speech as an input by using Google's Cloud Speech API. The output English text of Google Speech API is classified by using training data with the help of Naïve Bayesian Classifier to meet user desired text (i.e. contact name). Then the system calculates the similarity scores of the classified contact name, the transliterated name and English contact name in the contact list according to Jaro-Winkler similarity method. And it compares the similarity scores of two strings (string1 is the classified name and string2 is the combination of transliterated name and English contact name) and it chooses the highest similarity score. If the highest similarity score is greater than or equal to threshold value 8.4, the system will make phone call to that contact name. The Fig 1 shows architecture of the system.

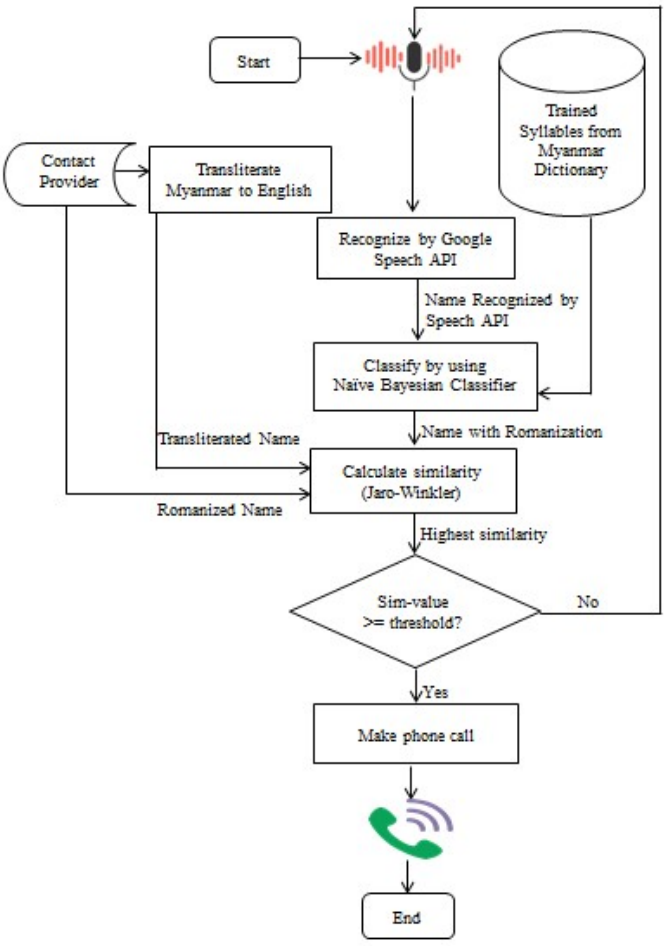

Fig.1. Architecture of the system

IV. Myanmar to English Transliteration

In this step, to get the contact names spelled with English script the system transliterates the contact names written with Myanmar script with the following steps as shown in Figure 2.

- Normalization
- Syllable Segmentation
- Transliteration with Hybrid Approach

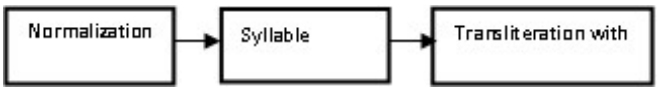

Fig.2. Process of Myanmar to English Transliteration

A. Text Normalization

Myanmar words can be classified into standard words (i.e., words with standard syllable structure) and irregular words (i.e., words with abbreviated characters or words written in special traditional writing forms etc.). Since these irregular words are written in different forms, it complicates the syllabication process. To identify correct syllable boundaries in the given text, it must have standard syllable structure. Therefore the irregular Myanmar words are needed to normalize to outperform the syllable segmentation result. Table 1 shows some examples for normalization of irregular words.

Table 1 Normalization of Myanmar Irregular Words

Irregular Word	Normalization Output
တက္ကသိုလ	တက်ကသိုလ်[te' ka- tho]
အင်္ဂါ	အင်ဂါ[in ga]
ယောကျ်ားး	ယောက်ကျား[jau' kya:]
ပဉ္ဇနာ	ပဉ်သနာ[pja' tha- na]

B. Syllable Segmentation

Myanmar is a syllabic script and syllable is a smallest unit in Myanmar language. The combination of one or more characters but not more than eight characters will become one syllable; combination of one or more syllables becomes one word. Syllable segmentation is the process of determining the syllable boundaries in a sentence or a document. In this system, syllable segmentation is done based on a Myanmar word segmentation tool [6]. For example – သန်တာချယ်ရီ → သန်[than] + တာ[da] + ချယ်[che] + ရီ[ji].

C. Transliteration with Hybrid Approach

The system accepts the syllable segments and transliterates them with the help of rules from BGN/PCGN 1970 + KNAB modification 2002 and Transliteration: KNAB 2003 Agreement to accurate our measurement standard [6, 7]. Although these rules solve almost the transliteration problems, they cannot solve in special case such as the spelling of loan word "ချယ်ရီ[cherry]". Therefore the system applies direct mapping transliteration approach to solve this special case. The system uses a dictionary for direct mapping transliteration approach. Table 2 and Table 3 describe transliterated output of rule based approach and hybrid approach respectively.

Table 2 Transliterated Output with Only Rule Based Approach

Splitting Word	သန်[than]	တာ[da]	ချယ်[che]	ရီ[ji]
Transliterated Word	than	dar	Chal	yi

Table 3 Transliterated Outputs with Hybrid Approach

Splitting Word	သန်[than]	တာ[da]	ချယ်[che]	ရီ[ji]
Transliterated Word	than	dar	Cherry	

V. Training Data

In this system, training data is required to get user desired contact name because speech API can only produce output text that is most similar pronunciation of input Myanmar proper name. According to Myanmar Orthography, there are 1864 unique Myanmar syllables in which some syllables are used to spell proper name but some are not (e.g. "ကိစ်[kei']", "ဂိဇ်[gei']", "ဇောဋ်[zau']", etc.).These syllables are trained with the help of Google speech API in this system. Before speech to text conversion of speech API, Myanmar to English transliteration for the above Myanmar syllables is required because the training data, English script output of speech API, is stored in the database. According to BGN/PCGN 1970 + KNAB modification 2002 and Transliteration: KNAB 2003 Agreement for the Transliteration of Myanmar into English, Myanmar syllables that are similar pronunciation, use the same rules to transliterate English scripts. Table 4 shows example of these transliteration rules.

Table 4 Example of Myanmar to English Transliteration

Myanmar Script	Romanization
ဒ,ဓ, ဎ, ဍ	D
ယ,ရ	Y
ဗ,ဘ	B
ကျမ်း, ကမြိုး	Kyan
ညီ,ညီး ညင်	Nyi
ဗန်,ဗမ်,ဗဏ်	Ban
ယင်,ရင်,ယဉ်, ယာဉ်	Yin
လတ်,လပ်,လာဘ်	Lat
ဒန်,ဒံ,ဒဏ်,ဓမ်	dan

Although there are 1864 unique Myanmar syllables, the training dataset used in this system contains only 878 syllables in which each syllable is defined as one class label. In other word, the dataset holds 878 unique class labels in which each class label is trained 5 times. After each training time for one syllable, the number and value of resulted training data are different according to the accent of input speech. Therefore there are totally 13,748 training data in this system. For example "kyaw" class label (syllable) is trained with Google's speech Recognition API as the following Table 5.

Table 5 Training Data of Class Label "kyaw"

Output of Speech API	Class Label
joel	kyaw
jole	kyaw
jo	kyaw
jojo	kyaw
joe	kyaw
joh	kyaw
joel	kyaw
jo	kyaw
jojo	kyaw
joe	kyaw
dole	kyaw
jo	kyaw

jojo	kyaw
joel	kyaw
joe	kyaw

This system trains the syllables with unigram as described above. The combinations of syllables frequently used in Myanmar proper names can be trained with bigrams. The Table 6 describes examples of training data of Myanmar syllables with bigrams. If syllables are trained with bigrams, contact name classified by Naïve Bayes and user's desired contact names are more similar than training with unigram. The similarity values of unigram and bigrams are not easy to train all possible bigram combinations because there are too many possible bigram combinations. But, this system is a mobile phone based application. It stores training data in the client and memory allocation is very important. Therefore the system trains syllables with unigram and it uses threshold value for smoothing of system output.

Table 6 Training Data of Bigrams for Class Label "su khin"

Output of Speech API	Class Label
soup can	sukhin
suitcase	sukhin
soup kitchen	sukhin
suki	sukhin
soup king	sukhin
soup can	sukhin
suit cane	sukhin
sukan	sukhin
sukin	sukhin
sue cain	sukhin
soup can	sukhin
soup kitchen	sukhin
suit can	sukhin
soupcon	sukhin
soup canned	sukhin

VI. Google's Speech Recognition Engine or API

The main requirement of every speech to text conversion system is a database which will compare pitches with frequencies. If we develop the system which will convert the speech into text that is for any user, it is very difficult job because the frequency of a user is different from that of other user. If the system is global hence creating the speech database for the mobile user is very much difficult because comparison of sound frequency and pitch for millions and billions mobile users is difficult again. To solve this issue of speech recognition (ASR) systems on mobile devices, Google Speech recognition engine or API which use a huge large speech database regarding possible different pitches and frequencies of a person's voice, is used to recognize user speech in this system. The Cloud Speech-to-Text uses a speech recognition engine that can understand one of a wide variety of languages (e.g. English, Mandarin, Chinese, Japanese and so on). This system uses English language which is one of the languages supported by Google Speech API. In some cases (e.g. Myanmar proper name), although the API returns output text to the client, these texts are not correct but similar pronunciation. The following Figure 2 is an example for the output text of Google's speech API.

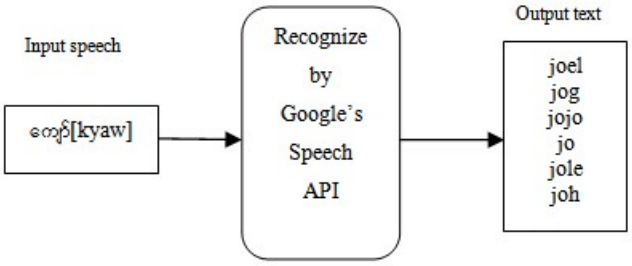

Fig 2 Output of Google's Speech API for Myanmar Proper Name

VII. Classification with Naïve Bayesian Classifier

The Google's speech API cannot correctly recognize Myanmar proper name but can produce English word that is similar pronunciation of that input Myanmar proper name. For example, if speech input is "yee/wai/khaing", API recognizes it as "sea/red/khine" this is not exact output. To get the correct contact name, this system requires some training to be done on the data input. Therefore Naïve Bayesian Classifier will classify to get the correct name for output name recognized with speech API by using the above described training data. To calculate the probability of a hypothesis(C) being true, given the prior knowledge(X), Bayes' Theorem is used as follows:

$$P(C|X)=(P(X|C)P(C))/(P(X))$$

The following Figure 3 is an example for classifying the speech API output text by using Naïve Bayesian Classifier.

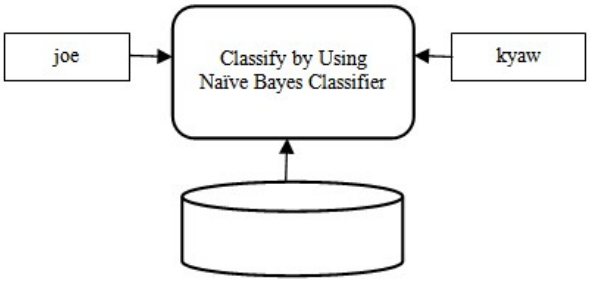

Fig. 3 Classified Text of Naïve Bayes Classifier

VIII. Making Phone Call

In this step, the system compares the similarity scores calculated by Jaro-Winker method and chooses the highest similarity score. If the highest similarity value is greater than or equal threshold value 8.4, the system will make phone call to the contact name. Decision of making phone call to the contact name depends on threshold value. As a result the system needs a good threshold value to make phone call to the user desired contact name. Therefore the system was tested with 463 contact names (311 non-nasalized contact names plus 152 nasalized contact names) to get good threshold value. These contact names are categorized into nasalized and non-nasalized names based on Myanmar vowel sounds. From threshold value 8.2 to 8.6 are set to know which threshold value is optimized for this system as described in Table 7. According to Table 7, if threshold value 8.5 is set, the success rate of correct phone call to non-nasalized contact name is below 80%, if threshold value 8.3 is set, the success rate of correct phone call to non-nasalized contact name is equal to that of correct phone call to non-nasalized contact name when threshold value is 8.4, but the success rates of incorrectly phone call to nasalized contact name are significantly different. Therefore the system specifies the good value of threshold value is 8.4.

Threshold Value	Type of Contact Name	Correct Phone Call	Incorrect Phone Call	Not Phone Call
8.2	non_nasalize	256(82.32%)	39(12.54%)	16(5.14%)
	nasalize	75(49.34%)	52(34.21%)	25(16.45%)
8.3	non_nasalize	255(81.99%)	36(11.58%)	19(6.11%)
	nasalize	75(49.34%)	43(28.29%)	33(21.71%)
8.4	non_nasalize	255(81.99%)	30(9.65%)	25(8.04%)
	nasalize	75(49.34%)	36(23.68%)	41(26.95%)
8.5	non_nasalize	243(78.14%)	26(8.36%)	29(9.32%)
	nasalize	71(46.71%)	30(19.74%)	47(30.92%)
8.6	non_nasalize	232(74.60%)	23(7.4%)	32(10.29%)
	nasalize	65(42.76%)	25(16.45%)	52(34.21%)

Table 7 Success Rates of Each Threshold Value

IX. Evaluation of the System

This section reports the experimental results of Myanmar to English Transliteration, speech to text (STT) conversion of Google's speech API and making phone call processes. The performance of each process applied in this system is measured with success rate. A success rate is one of many metrics used for measuring/quantifying usability and is the simplest accuracy measurement method. To get the value of success rate for each process, the following formula is used.

$$Success\ rate = \frac{no\ of\ correct\ data}{no\ of\ total\ testing\ data} \times 100$$

A. Experimental Result of Myanmar to English Transliteration

Two types of Myanmar words (standard and irregular word) have been discussed in Section 4.2. Since the words are written in different orthographic ways, it may cause problem to the transliteration system. Therefore this system has been tested with 348 distinct Myanmar proper names for transliteration process performance as shown in Table 8.

Table 8 Performance of Myanmar to English Transliteration

No	Type of Word	No. of Contact Name	No. of Correctly Transliterated Name	% of Correct Translitern for Each Word Type
1.	Standard word	286	279	97%
2.	Irregular word	62	57	91%
TOTAL		348	336	96%

B. Consequences of Google's Speech API

Google's speech API can recognize 120 languages and variants to support global user base. The Google's speech API cannot correctly recognize some languages such as Myanmar Language but it can produce output text which pronunciation is similar to the input Myanmar word. Therefore the consonants in each consonant cluster have been analyzed to measure the performance of Google's Speech API. The accuracy of Google's cloud speech API is shown in Table 9

Table 9 Accuracy of Google's Speech API for Each Consonant Cluster

No	Type of Consonant Cluster	Usage Level	% of Correct STT Conversion of API
1	$0C_20$	Highest	87.5%
		Medium	81.82%
		Lowest	83.33%
2	$0C_2C_3$	Highest	81.82%
		Lowest	66.67%
3	C_1C_20	Highest	66.67%
		Lowest	50%
4	$C_1C_2C_3$		25%
TOTAL			71.67%

C. Experimental Result of Overall System

The contact name consists of one or more syllables. The syllable may be only vowel or combination of vowel and consonant. There are different types of consonants in Myanmar language already explained in Section 2.15. The system has been tested with contact names based on Myanmar consonant types. The following Table 10 shows the performance of the system for each consonant type.

Table 10: Performance of the Overall System

No	Type of Consot	No. of Contact Name	No. of Making Phone Call to Contact Name	% of Correctly Making Phone Call
1	Nasal	548	366	66.78%
2	Stop	563	409	72.64%
3	Fricative	523	354	67.68%
4	Affricate	186	143	76.88%
5	Central Approximant	280	212	75.71%
6	Lateral Approximant	183	136	74.31%

According to the above Table 10, performance of the system especially decreases when a contact name contains either nasal or fricative consonant. Therefore the contact names are categorized into two groups: contact name that contains nasal or fricative consonants and contact name that does not contain both nasal and fricative consonants and the system was analyzed with these two groups. The following Table 11 concludes the accuracy result of the system.

Table 11: Accuracy Result of the Overall System

No	Type of Consonant	No. of Contact Name	No. of Making Phone Call to Contact Name	% of Correctly Making Phone Call
1	With nasal or fricative	781	533	68.24%
2	Without nasal and fricative	119	108	90.75%
Total		900	641	71.22%

X. Result Discussion

The performance of Google's speech API, transliteration process and overall system have been analyzed. Firstly, Myanmar to English transliteration process was tested with 348Myanmar proper names. It observed that only 12 proper names out of 348 proper names are erroneous. Therefore it received 96% of overall accuracy covering both types of standard and irregular words. By doing error analysis, the errors are caused by those words with different styles in written and spoken format (e.g. "ပုသိမ်ကြီး[pa- thein gyi:]", "မစေ့ပယ်ညို [may sa- be nyo]"). Secondly, Google's speech API was tested with different kinds of Myanmar consonant clusters. There are four consonant clusters (0C20, 0C2C3, C1C20 and C1C2C3). The success rate of Google's speech API for Myanmar consonants is 71.67%. As a result, it is found that the speech API cannot correctly recognize the consonants which are not widely used such as "ဿ [nhwa.]", "လ္ဟ [lhwa.]", "မျှ [mjha.]" and so on. In this system, the performance of making phone call process totally depends on speech API. The Myanmar contact names contain one or more syllables. Each syllable consists of at least one vowel or combination of consonant and vowel. The performance of the overall system was analyzed with 900 testing contact names. These contact names were categorized into different. We conclude that the success rate values of these contact names are from 66.78 % to 90.75% respectively. As a result, it was examined that there are contact name recognition errors when all syllables of the contact name are vowels (e.g. "အေးအေးအောင်[ei: ei: aun]", "အိအိ[ei ei]"), contact name contains nasal or fricative consonants only (e.g. "မင်္ဂြမြတ်မွန်[mjin. mja' mun]", "ဆုဇော်ဇော် ဇင်[hsu. zo zo zin]") and combination of nasal and fricative consonants (e.g. "ငုဝါစိုး [ngu. wa so:]", "ညီဇေယျာလိုင် [nyi zei ja lhain]").

XI. Conclusion

Natural language processing techniques are becoming widely popular scientific research areas as well as Information Technology industry. Language technology together with Information Technology can enhance the lives of people with different capabilities. This system implements voice command mobile phone dialer application. The strength of the system is that it can make phone call to the contact name written in either English script or Myanmar scripts.

REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955.

[2] B. R. Reddy, E. Mahender, "Speech to Text Conversion using Android Platform" , International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com Vol. 3, Issue 1, January - February 2013, pp.253-258.

[3] S. Dash, Master of Engineering , Department of Computer Science & Engineering, Thapar University, Patiala, Punjab, India, "Kuiz Apps Using Voice Detection in Android Platform", International Journal of Computer Engineering and Applications,

[4] Ms. A. Jadhav, Prof. A. Patil, "Android Speech to Text Converter for SMS Application", IOSR Journal of Engineering Mar. 2012, Vol. 2(3) pp: 420-423.

[5] H. D. J. Jeong, S. K.Ye, J. Lim, I. You, W. S. Hyun, Department of Computer Software, Korean Bible University Seoul, South Korea, "A Computer Remote Control System Based on Speech Recognition Technologies of Mobile Devices and Wireless Communication Technologies", IEEE Conference Publication,2013,page no. 595-600.

[6] A. M. Mon, S. L. Phyue, M. M. Thein, S. S. Htay, T. T. Win, "Analysis of Myanmar Word Boundary and Segmentation by Using Statistical Approach", International Conference on Advanced Computer Theory and Engineering (ICACTE), Vol.5, Augest 2010.

[7] Romanization System for Burmese: BCG/PCGN 1970 Agreement, Office of the Superintendent, Government Printing, Rangoon, Burma.

[8] T. H. Hlaing, Yoshiki MIKAMI, "Automatic Syllabification of Myanmar Texts using Finite State Transducer", International Journal on Advances in ICT for Emerging Regions, Volume 6, Number 2 2013.